



09/886 .899

08.25.03



INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

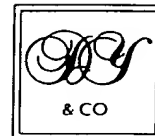
Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

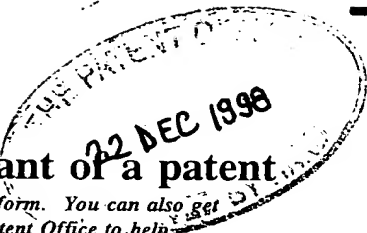
Andrew Gersey

Dated

27 June 2003



Z3DEC98 E414045-16 D02246
P01/7700 0.00 - 9828383.1



Request for a grant of a patent

See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference P006079GB ATM

2. Patent application number
(The Patent Office will fill in this part) **9828383.1**

3. Full name, address and postcode of the or of each applicant
(underline all surnames) Medical Research Council
20 Park Crescent
London
W1N 4AL
United Kingdom

Patents ADP number (if you know it) 596007001

If the applicant is a corporate body, give the country/state of its incorporation United Kingdom

4. Title of the invention Cell Lineage Markers

5. Name of your agent (if you have one) D YOUNG & CO

"Address for service" in the United Kingdom to which all correspondence should be sent
(including the postcode) 21 NEW FETTER LANE
LONDON
EC4A 1DA

Patents ADP number (if you have one) 59006

6. If you are declaring priority from one or more earlier patent applications, give the country and date of filing of the or each of these earlier applications and (if you know it) the or each application number	Country	Priority application number (if you know it)	Date of filing (day/month/year)
---	---------	---	------------------------------------

7. If this application is divided or otherwise derived from an earlier UK application, give the number and filing date of the earlier application	Number of earlier application	Date of filing (day/month/year)
---	-------------------------------	------------------------------------

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if:
a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

YES

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form 0

Description 45

Claims(s) 3

Abstract 1

Drawing(s) 0

10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents

Statement of inventorship and right to grant of a patent (Patents Form 7/77)

Request for preliminary examination and search (Patents Form 9/77)

Request for substantive examination (Patents Form 10/77)

Any other documents (please specify)

11.

I/We request the grant of a patent on the basis of this application.

Signature

Date

D YOUNG & CO
Agents for the Applicants

22 12 98

12. Name and daytime telephone number of the person to contact in the United Kingdom

Dr A Maschio

01703 634816

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505

b) Write your answers in capital letters using black ink or you may type them

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

CELL LINEAGE MARKERS

The present invention relates to a method of marking, selecting and generating
5 committed or partially committed cell lineages from tissues. In particular, the invention
relates to the use of the *Sox* genes for the selection or generation of various specified
cell types.

SOX proteins constitute a family of transcription factors related to the mammalian testis
10 determining factor SRY through homology within their HMG box DNA binding
domains. In DNA binding studies, SOX proteins exhibit sequence Specific binding;
however, unlike most transcription factors, binding occurs in the minor groove
resulting in the induction of a dramatic bend within the DNA helix. Although SOX
proteins can induce transcription of reporter constructs in vitro and possess activation
15 domains, transcriptional activation by these factors appears to be context dependent. In
other words members of this family finally seem to act in conjunction with other proteins.
Therefore, SOX proteins display properties of both classical transcription factors and
architectural components of chromatin (reviewed by Pevny and Lovell-Badge, 1997).

20 Members of the *Sox* gene family are expressed in a variety of embryonic and adult
tissues, where they appear to be responsible for the development and/or elaboration of
particular cell lineages. *Sry* is transiently expressed in the precursor Sertoli cells of the
XY genital ridge and is responsible for triggering development of the male phenotype
(reviewed by Lovell-Badge and Hacker, 1995). Thus, the lack of *Sry* results in XY
25 females and XX males. *Sox9* is expressed in immature chondrocytes and male gonads;
mutations in the human SOX9 gene are associated with Campomelic Dysplasia, a
human skeletal malformation syndrome, and XY female sex reversal. *Sox4* is
expressed in many tissues and a null mutation of the gene in mouse results in the
absence of mature B cells and heart malformations. *Xsox17* genes are involved in
30 endoderm formation in *Xenopus* embryos. These functional analyses Suggest that *Sox*
genes function in cell fate decisions in diverse developmental pathways.

A subfamily of *Sox* genes, that includes *Sox1*, *Sox2* and *Sox3*, shows expression profiles during vertebrate embryogenesis that suggest the genes could function in the control of cell fate decisions within the early developing nervous system. *Sox2* and *Sox3* begin to
5 be expressed at preimplantation and epiblast stages respectively, and are then restricted to the neuroepithelium. *Sox1* appears only at around the stage of neural induction. Related to *Sox 1- 3* are *Sox 14*, *Sox 19* and *Sox 70D*, which are also found at various stages during the development of neural tissues. A number of other *Sox* genes, and their tissue distributions, are known. See Pevny & Lovell-Badge, (1997) Curr. Opin.
10 Genet. Dev., 7:338-344, incorporated herein by reference.

The molecular mechanisms controlling induction and determination of tissue development during embryogenesis have begun to be elucidated. The identification by cellular and biochemical methods, of secreted molecules involved in the development of
15 cell fate illustrates the important role of the environment in specifying cell identity. In addition, a number of transcription factors have been isolated which play important roles in the specification and differentiation of neural cell lineages. For example, the characterisation of vertebrate homologues of *Drosophila* proneural and neurogenic genes, which control neural specification in the fly, has revealed analogous molecular
20 mechanisms in vertebrate neural cell fate determination and differentiation. In an *Drosophila*, the expression of basic helix-loop-helix transcription factors of the AS-C complex confirms neural potential on groups of ectodermal cells. Misexpression of transcription factors involved in cell fate determination is observed to cause abnormalities in development.

25

In our copending international patent application PCT/GB98/01862, filed 25th June 1998, we describe the use of the *Sox 1* gene and SOX 1 polypeptide in inducing commitment to the neural pathway in pluripotent embryonal stem cells, and in identifying neurally-committed cells. The possibility, or otherwise, of employing other
30 SOX proteins is not discussed.

Summary of the Invention

5 In accordance with the present invention, it has been found that *Sox* gene expression correlates in general with specific stages during embryogenesis. Moreover, it has been determined that the expression of *Sox* genes may be used, as directed herein, to induce or select pluripotent cells which are at least partially committed to a given developmental pathway.

10 According to a first aspect of the present invention, there is provided a method for isolating a pluripotent cell which is at least partially committed to a given developmental pathway, comprising the steps of :

- a) selecting a population of pluripotent cells;
- b) sorting the cells according to *Sox* gene expression; and
- 15 c) isolating those cells which express a given *Sox* gene.

As set forth in the following description, the *Sox* genes, which encodes SOX proteins, are responsible for the specification of a variety of proliferating cells which are not yet totally committed, as well as acting as a marker for such cells. Expression of *Sox*
20 genes is responsible for the generation of specific pluripotent cell lineages, which *in vivo* or *in vitro* are capable of differentiating into the many different cells which belong to a given developmental line.

As used herein, a "pluripotent cell" is a cell which may be induced to differentiate, *in vivo* or *in vitro*, into at least two different cell types. These cell types may themselves
25 by pluripotent, and capable of differentiating in turn into further cell types, or they may be terminally differentiated, that is incapable of differentiating beyond their actual state. Pluripotent cells include totipotent cells, which are capable of differentiating along any chosen developmental pathway. For example, embryonal stem cells (Thomson *et al.*,
30 (1998) Science 282:1145-1147) are totipotent stem cells. Pluripotent cells also include

other, tissue-specific stem cells, such as neuronal stem cells, neuroectodermal cells, ectodermal cells and endodermal cells, for example gut endodermal cells.

5 “Developmental pathway” refers to a common cell fate which can be traced from a particular precursor cell. Thus, for example, the neuronal pathway defines those cells which developing from the neural plate, give rise to all the neural cells and ganglia of an adult organism. They can alternatively be defined as cells of the “neural lineage”.

10 A “partially committed” cell is a cell type which is no longer totipotent but remains pluripotent. For example, neuroectodermal cells are capable of giving rise to any cell type in the CNS or PNS, yet are not able to give rise to endodermal tissues.

15 Pluripotent cells may be “selected” by any one or more of a variety of means, and the term includes dissection of tissue types from developing embryos, isolation or generation of pluripotent, including totipotent, cells *in vivo* or *in vitro*. Preferably, the term refers to the isolation of one class of pluripotent cells from one or more other cell types. In the context of the present invention, this allows grater precision in selection using *Sox* genes because, as a result of their widespread expression, particular *Sox* genes cannot be generally stated to be exclusively associated with any one tissue.
20 Thus, preselection of possible tissue types allows *Sox* gene expression to be used to accurately identify a desired cell lineage from a remaining cell population.

25 Cells can be sorted by affinity techniques, or by cell sorting (such as fluorescence-activated cell sorting) where they are labelled with a suitable label, such as a fluorophore conjugated to or part of, for example, an antisense nucleic acid molecule or an immunoglobulin. As used herein, “sorting” refers to the at least partial physical separation of a first cell type from a second.

30 “Isolating” cells refers to removing at least one component from a mixture in which the cells were previously associated. In the context of the present invention, “isolating” preferably refers to removal of at least one cell type from a mixed population of cells.

Preferably, "isolating" can refer to the enrichment of a population of cells for a desired cell type. Preferably, "isolating" refers to substantial purification such that there is only a single cell type present in the final population.

5 According to a second aspect of the invention, cells can be actively sorted from other cell types by detecting the expression of SOX polypeptides *in vivo* using a reporter system. Thus, for example, the invention provides a method for isolating a desired cell type from a population of cells, comprising the steps of:

- (a) transfecting the population of cells with a genetic construct comprising a
10 coding sequence encoding a detectable marker operatively linked to *Sox* control regions;
- (b) detecting the cells which express the selectable marker; and
- (c) sorting the cells which express the selectable marker from the population of cells.

15 The selectable marker may be any selectable entity, but is preferably a fluorescent or luminescent marker which may be detected and sorted by automated cell sorting approaches. For example, the marker may be GFP or luciferase. Other useful markers include those which are expressed in the cell membrane, thus facilitating cell sorting by affinity means.

20

Sox control sequences are control sequences derived from *Sox* genes and which regulate the expression of SOX polypeptides. *Sox* control sequences are known in the art, as further described below.

25 According to a further aspect of the invention, cells can be actively sorted from other cell types by detecting the expression of SOX polypeptides *in vivo* using a reporter system which is itself responsive to *Sox* gene expression. Thus, for example, the invention provides a method for isolating a desired cell type from a population of cells, comprising the steps of:

(a) transfecting the population of cells with a genetic construct comprising a coding sequence encoding a detectable marker operatively linked to control regions sensitive to modulation by a SOX polypeptide;

(b) detecting the cells which express the selectable marker; and

5 (c) sorting the cells which express the selectable marker from the population of cells.

The genetic construct according to the invention may comprise any promoter and enhancer elements as required, so long as the overall control remains sensitive to a
10 SOX polypeptide; in other words, no expression of the marker coding sequence should take place in the absence of the desired SOX protein. The regulatory sequences of responsive to SOX polypeptides are known in the art and have been described in the literature cited herein and incorporated herein by reference; at least, however, the
15 construct of the invention will comprise a SOX binding site. Preferably, the natural SOX-responsive control elements are used in their entirety; however, other promoter and enhancer elements may be substituted where they remain under the influence of SOX expression.

The selectable marker will only be expressed in desired cell types because only these
20 cells express the relevant SOX polypeptide, which is required for transcription from the Sox control sequences. Preferably, therefore, the expression means used to express the selectable marker is not leaky and express a minimal amount of the marker in the absence of the SOX polypeptide. Techniques for transforming cells with coding genetic constructs according to the invention, detecting the marker and sorting cells accordingly
25 are known in the art.

The present invention, in a still further aspect, provides the use of Sox coding sequences to transform precursor cells and thereby differentiate desired partially committed cells therefrom. Accordingly, there is provided a method for differentiating
30 partially committed cell from a pluripotent precursor cell, comprising the steps of:

(a) transforming the pluripotent precursor cell with a genetic construct comprising a *Sox* coding sequence operatively linked to a suitable control sequences; and

5 (b) culturing the cells so as to allow expression of the *Sox* coding sequence, thereby inducing the cell to differentiate.

Suitable control sequences for use in the letter aspect of the invention are known in the art and may include inducible or constitutive control sequences. Inducible control sequences have the advantage that *Sox* gene expression may be switched off when
10 desired, for example once the cell is to be differentiated into another neural cell. Moreover, once the expression of exogenous *Sox* gene has been switched off, successfully differentiated neuroblasts may be identified by virtue of the continued expression of the endogenous *Sox* gene.

15 Precursor cells may be, for example, ES cells, such as human ES cells and cells with similar pluripotent properties derived from germ cells (EG cells). More specific pluripotent precursors or direct precursors of any desired cell lineage may also be employed.

20 Cells obtained according to the invention may be employed in a number of ways. Of course, the expression of *Sox* genes has important implications for the study of embryonal differentiation; the generation and selection of specific cell lineages will provide material for basic research.

25 Moreover, the invention has medical and diagnostic applications. The detection of *Sox* expressing cells is important in clinical neurology and in diagnosing and treating cancers of the nervous system. Accordingly, the invention provides a method for detecting the presence of a neuroblast as described above for diagnostic purposes.

30 Stem cells are also useful for the treatment of disorders of any given tissue, particularly for the treatment of neurological disorders and especially for repair of accidentally

induced trauma in the CNS or for the correction of congenital or pathological diseases of the CNS.

Moreover, in applications involving somatic gene therapy designed to correct a genetic defect, the removal, treatment and replacement of pluripotent cells which are actively dividing has clear advantages, providing a constant source of modified neural cells to permanently treat the targeted defect. *Sox* control sequences may be used specifically to direct transgene expression in specified cells where this is desired. Moreover, gene expression can be directed to terminally differentiated cell types derived from pluripotent cells by the use of other control sequences, such as NF-1 control sequences which direct expression of NF-1 in mature neurons *in vivo*.

A significant advantage of the methods described herein is that a patient in need of treatment can act as a self-donor. In other words, cells may be isolated from the patient and either sorted to extract desired cell types, or treated in order to differentiate the required cells as described, from specific or general precursors.

Detailed Description of the Invention

The present invention is directed to methods for isolating, or producing, cells of any desired lineage. The expression of *Sox* genes is associated with a wide variety of cell types. Table 1 is a non-exhaustive list of known *Sox* genes, and shows the cell lineages with which they are associated *in vivo*.

The temporal and tissue-specific expression patterns of *Sox* genes are the subject of study by many groups, and in many cases such patterns are well mapped. For example, *Sox1* expression appears to be limited to the neural plate and in induction of lens-associated gene expression in the eye. *Sox2* is more widespread in its expression patterns, being expressed widely in the preimplantation embryo, and effectively defining the totipotent lineage. During gastrulation it is turned off in the mesoderm, but remains active in prospective neuroectoderm.

Table 1. The SOX family.

Gene	Orthologue	Mapping	Expression
Group A			
SRY	Human, rodent, marsupial	Human Y	Genital ridge and testis
Group B			
SOX1	Human, mouse, chicken	Human 8	Sox1: embryonic CNS, lens
SOX2	Human, mouse, chicken	Human 3	Sox2: ICM, primitive ectoderm, CNS, PNS, embryonic gut, endoderm; cSox2: embryonic CNS and PNS
SOX3	Human, mouse, chicken	Human X	Sox3 and cSox3: embryonic CNS
Sox14	Mouse	-	-
ZfSox19	Zebrafish	-	Embryonic CNS
Sox15	Mouse	-	-
Dichaete (Sox70D, fish-hook)	Drosophila	70D1-2	Entire trunk of syncytial blastoderm; seven irregular stripes in cellular blastoderm; ventral and cephalic neuroectoderm
Group C			
SOX4	Human, mouse	Human 6	Embryonic heart and spinal cord, adult pre-B and pre-T cells
SOX11	Human, mouse, chicken	Human 2	cSox11: embryonic CNS,

SOX12	Human, mouse	–	post-mitotic neurons
SOX20	Human	Human 17	–
Group D			
SOX5	Human, mouse	Human 6	Sox5: adult testis
SOX6	Human, mouse	–	Sox6: embryonic CNS, adult testis
Group E			
SOX8	Human, mouse	–	–
SOX9	Human, mouse, chicken	Human 17, Mouse 2	Chondrocyte, genital ridge and adult testis, CNS, notochord
SOX10	Human, mouse	–	–
Group F			
Sox-7	Mouse	–	–
Sox-17	Mouse	–	Testis
Sox-18	Mouse	Mouse 2	Heart, lung, spleen, skeletal muscle, liver and brain of adult ICM, inner cell mass; PNS, peripheral nervous system.

At this stage of differentiation, therefore, *Sox2* has become a marker for cells committed to the neural lineage, but still capable of differentiation into a variety of cell types within that lineage. However, it is also expressed in gut endoderm and ectodermal lineages which give rise to eye, olfactory, ear and hair follicle tissues.

5

Sox3 is expressed throughout the ectoderm before gastrulation, and then becomes largely restricted to the neuroectoderm, as with *Sox1* and *Sox2*. Although not as widely expressed as *Sox2*, it does retain expression at some mesodermal locations.

10 *Sox4* is expressed in embryonic heart and spinal chord, and adult pre-B and T lymphocytes.

The method of the invention does not require absolutely unique expression of a *Sox* gene in order to isolate partially committed pluripotent cells. The present invention provides that *Sox* genes in general are markers for the state of pluripotency, rather than
15 for any particular tissue. Accordingly, tissues or cell types may be sorted, for example by dissection of relevant tissues from embryos, or by induction of differentiation in cells in order to produce suitable cell populations; *Sox* gene expression may then be used to detect or induce a pluripotent cell type in the selected population of cells.

20

At least the following *Sox* genes are known; others may be isolated by homology searching. *Sox 21* (GenBank Accession No. AF107044); *Sox 14* (GenBank Accession No. 107043); *Sox 13* (GenBank Accession No. AB104474); *Sox 10* (GenBank Accession No. AJ001183); *Sox 22* (GenBank Accession No. U35612); *Sox 18*
25 (GenBank Accession No. L35032); *Sox 11* (GenBank Accession No. U23752); *Sox 1* (GenBank Accession No. Y13436); *Sox 2* (GenBank Accession No. z31560 and U12532); *Sox 3* (GenBank Accession No. X94125); *Sox 4* (GenBank Accession No. X70683); *Sox 5* (GenBank Accession No. S83306); *Sox 6* (GenBank Accession No. U32614); *Sox 7* (GenBank Accession No. AI15903/P40646); *Sox 9* (GenBank
30 Accession No. S74504/5/6); *Sox 12* (GenBank Accession No. U70442); *Sox 13* (GenBank Accession No. AB006329); *Sox 15* (GenBank Accession No. AB104474);

Sox 16 (GenBank Accession No. L29084); *Sox 17* (GenBank Accession No. D49473);
Sox 19 (GenBank Accession No. X98368); *Sox 22* (GenBank Accession No. U35612).

Sox genes are divisible into subfamilies, as indicated in Table 1. *Sox 1, 2* and *3* belong
5 to a single subfamily, Group B. Expression of this *Sox* gene subfamily has been
evolutionarily conserved. The *Drosophila* (Nambu and Nambu 1996; Russel et al.,
1996) zebrafish (Vriz et al., 1996) and avian (Unwanogho et al., 1995; Streit et al.,
1997; Rex et al., 1997) putative orthologues of *Sox1*, *Sox2* and *Sox3* all show
expression throughout the neural primordium. Thus, this subfamily of *Sox* genes
10 represents a novel group of transcription factors which can serve as general early
neuroepithelial markers. Similar conservation is observable amongst other subfamilies,
known as Groups C - F.

In general, *Sox* proteins and genes as referred to herein may be derived from any
15 source, preferably from a mammalian source such as human or mouse, but also from
other sources, such as fish and insect.

A number of *Sox* gene sequences are known in the art and provided under the GenBank
accession numbers given above. Other *Sox* sequences may be isolated, for example
20 from genomic or cDNA libraries, by conventional techniques. The sequences provided
herein may be used as probes, or to prepare antibodies or other molecules capable of
recognising specific polypeptides. Preferably, the sequences used as probes are
substantially homologous to the sequences provided herein.

25 "Substantial homology", where homology indicates sequence identity, means more than
40% sequence identity, preferably more than 45% sequence identity and most
preferably a sequence identity of 50% or more. Advantageously, the sequence identity
may be up to about 90 or 95%.

30 Sequence homology (or identity) may be determined using any suitable homology
algorithm, using for example default parameters. Advantageously, the BLAST

algorithm is employed, with parameters set to default values. The BLAST algorithm is described in detail at http://www.ncbi.nih.gov/BLAST/blast_help.html, which is incorporated herein by reference. The search parameters are defined as follows, and are advantageously set to the defined default parameters.

5

BLAST (Basic Local Alignment Search Tool) is the heuristic search algorithm employed by the programs `blastp`, `blastn`, `blastx`, `tblastn`, and `tblastx`; these programs ascribe significance to their findings using the statistical methods of Karlin and Altschul (1990, 1993) with a few enhancements. The BLAST programs were tailored for
10 sequence similarity searching, for example to identify homologues to a query sequence. The programs are not generally useful for motif-style searching. For a discussion of basic issues in similarity searching of sequence databases, see Altschul et al. (1994).

The five BLAST programs available at <http://www.ncbi.nlm.nih.gov/BLAST> perform
15 the following tasks:

blastp compares an amino acid query sequence against a protein sequence database;

blastn compares a nucleotide query sequence against a nucleotide sequence database;

20

blastx compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database;

tblastn compares a protein query sequence against a nucleotide sequence database
25 dynamically translated in all six reading frames (both strands).

tblastx compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

30 BLAST uses the following search parameters:

HISTOGRAM Display a histogram of scores for each search; default is yes. (See parameter H in the BLAST Manual).

5 DESCRIPTIONS Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 100 descriptions. (See parameter V in the manual page). See also EXPECT and CUTOFF.

10 ALIGNMENTS Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 50. If more database sequences than this happen to satisfy the statistical significance threshold for reporting (see EXPECT and CUTOFF below), only the matches ascribed the greatest statistical significance are reported. (See parameter B in the BLAST Manual).

15 EXPECT The statistical significance threshold for reporting matches against database sequences; the default value is 10, such that 10 matches are expected to be found merely by chance, according to the stochastic model of Karlin and Altschul (1990). If the statistical significance ascribed to a match is greater than the EXPECT threshold, the match will not be reported. Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Fractional values are acceptable. (See
20 parameter E in the BLAST Manual).

CUTOFF Cutoff score for reporting high-scoring segment pairs. The default value is calculated from the EXPECT value (see above). HSPs are reported for a database sequence only if the statistical significance ascribed to them is at least as high as would
25 be ascribed to a lone HSP having a score equal to the CUTOFF value. Higher CUTOFF values are more stringent, leading to fewer chance matches being reported. (See parameter S in the BLAST Manual). Typically, significance thresholds can be more intuitively managed using EXPECT.

30 MATRIX Specify an alternate scoring matrix for BLASTP, BLASTX, TBLASTN and TBLASTX. The default matrix is BLOSUM62 (Henikoff & Henikoff, 1992). The valid

alternative choices include: PAM40, PAM120, PAM250 and IDENTITY. No alternate scoring matrices are available for BLASTN; specifying the MATRIX directive in BLASTN requests returns an error response.

- 5 STRAND Restrict a TBLASTN search to just the top or bottom strand of the database sequences; or restrict a BLASTN, BLASTX or TBLASTX search to just reading frames on the top or bottom strand of the query sequence.

10 FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (Computers and Chemistry, 1993), or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (Computers and Chemistry, 1993), or, for BLASTN, by the DUST program of Tatusov and Lipman (in preparation). Filtering can eliminate statistically significant but biologically
15 uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

Low complexity sequence found by a filter program is substituted using the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNNN") and the letter "X" in protein
20 sequences (e.g., "XXXXXXXXXX"). Users may turn off filtering by using the "Filter" option on the "Advanced options for the BLAST server" page.

Filtering is only applied to the query sequence (or its translation products), not to
25 database sequences. Default filtering is DUST for BLASTN, SEG for other programs.

It is not unusual for nothing at all to be masked by SEG, XNU, or both, when applied to sequences in SWISS-PROT, so filtering should not be expected to always yield an effect. Furthermore, in some cases, sequences are masked in their entirety, indicating
30 that the statistical significance of any matches reported against the unfiltered query sequence should be suspect.

NCBI-gi Causes NCBI gi identifiers to be shown in the output, in addition to the accession and/or locus name.

- 5 Most preferably, sequence comparisons are conducted using the simple BLAST search algorithm provided at <http://www.ncbi.nlm.nih.gov/BLAST>.

Preferably, the invention makes use of fragments of Sox sequences. Fragments of the nucleic acid sequence of a few nucleotides in length, preferably 5 to 150 nucleotides in
10 length, are especially useful as probes.

Exemplary nucleic acids, including those of new Sox clones derived according to the invention can alternatively be characterised as those nucleotide sequences which encode a SOX protein and hybridise to the DNA sequences set forth above, or a selected
15 fragment of said DNA sequences. Preferred are such sequences encoding SOX polypeptides which hybridise under high-stringency conditions to the sequence set forth above.

Stringency of hybridisation refers to conditions under which polynucleic acids hybrids
20 are stable. Such conditions are evident to those of ordinary skill in the field. As known to those of skill in the art, the stability of hybrids is reflected in the melting temperature (T_m) of the hybrid which decreases approximately 1 to 1.5°C with every 1% decrease in sequence homology. In general, the stability of a hybrid is a function of sodium ion concentration and temperature. Typically, the hybridisation reaction is performed under
25 conditions of higher stringency, followed by washes of varying stringency.

As used herein, high stringency refers to conditions that permit hybridisation of only those nucleic acid sequences that form stable hybrids in 1 M Na⁺ at 65-68 °C. High stringency conditions can be provided, for example, by hybridisation in an aqueous
30 solution containing 6x SSC, 5x Denhardt's, 1 % SDS (sodium dodecyl sulphate), 0.1 Na⁺ pyrophosphate and 0.1 mg/ml denatured salmon sperm DNA as non specific

competitor. Following hybridisation, high stringency washing may be done in several steps, with a final wash (about 30 min) at the hybridisation temperature in 0.2 - 0.1x SSC, 0.1 % SDS.

- 5 Moderate stringency refers to conditions equivalent to hybridisation in the above described solution but at about 60-62°C. In that case the final wash is performed at the hybridisation temperature in 1x SSC, 0.1 % SDS.

- 10 Low stringency refers to conditions equivalent to hybridisation in the above described solution at about 50-52°C. In that case, the final wash is performed at the hybridisation temperature in 2x SSC, 0.1 % SDS.

- It is understood that these conditions may be adapted and duplicated using a variety of buffers, e.g. formamide-based buffers, and temperatures. Denhardt's solution and SSC
15 are well known to those of skill in the art as are other suitable hybridisation buffers (see, e.g. Sambrook, et al., eds. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York or Ausubel, et al., eds. (1990) *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.). Optimal hybridisation conditions have to be determined empirically, as the length and the GC
20 content of the hybridising pair also play a role.

- Advantageously, the invention moreover provides nucleic acid sequence which are capable of hybridising, under stringent conditions, to a fragment of a *Sox* as set forth above. Preferably, the fragment is between 15 and 50 bases in length.
25 Advantageously, it is about 25 bases in length.

- As will be appreciated by those skilled in the art, the redundancy of the genetic code allows the design of a large number of sequences encoding SOX polypeptides. Any of these sequences may be useful for expressing SOX polypeptides as described below.
30 An advantage of the use of a sequence encoding human SOX1 which is not the human Sox1 sequence is that the mRNA produced has a different sequence to that of the

endogenous SOX mRNA, and may thus be distinguished therefrom. Antisense oligonucleotides may be designed which are capable of selectively inhibiting the expression of either endogenous or exogenous *Sox* genes.

- 5 Given the guidance provided herein, nucleic acids encoding SOX polypeptides are obtainable according to methods well known in the art. For example, a nucleic acid encoding SOX polypeptides is obtainable by chemical synthesis, using polymerase chain reaction (PCR) or by screening a genomic library or a suitable cDNA library prepared from a source believed to express SOX polypeptides and to express it at a detectable
10 level.

Chemical methods for synthesis of a nucleic acid of interest are known in the art and include triester, phosphite, phosphoramidite and H-phosphonate methods, PCR and other autoprimer methods as well as oligonucleotide synthesis on solid supports. These
15 methods may be used if the entire nucleic acid sequence of the nucleic acid is known, or the sequence of the nucleic acid complementary to the coding strand is available. Alternatively, if the target amino acid sequence is known, one may infer potential nucleic acid sequences using known and preferred coding residues for each amino acid residue.

20

An alternative means to isolate genes encoding SOX polypeptides is to use PCR technology as described e.g. in section 14 of Sambrook et al., 1989. This method requires the use of oligonucleotide probes that will hybridise to *Sox* nucleic acid. Strategies for selection of oligonucleotides are described below.

25

Libraries are screened with probes or analytical tools designed to identify the gene of interest or the protein encoded by it. For cDNA expression libraries suitable means include monoclonal or polyclonal antibodies that recognise and specifically bind to SOX polypeptides; oligonucleotides of about 20 to 80 bases in length that encode known or
30 suspected *Sox* cDNA from the same or different species; and/or complementary or homologous cDNAs or fragments thereof that encode the same or a hybridising gene.

Appropriate probes for screening genomic DNA libraries include, but are not limited to oligonucleotides, cDNAs or fragments thereof that encode the same or hybridising DNA; and/or homologous genomic DNAs or fragments thereof.

- 5 A nucleic acid encoding SOX polypeptides may be isolated by screening suitable cDNA or genomic libraries under suitable hybridisation conditions with a probe, i.e. a nucleic acid disclosed herein including oligonucleotides derivable from the sequences set forth above. Suitable libraries are commercially available or can be prepared e.g. from cell lines, tissue samples, and the like.

10

As used herein, a probe is e.g. a single-stranded DNA or RNA that has a sequence of nucleotides that includes between 10 and 50, preferably between 15 and 30 and most preferably at least about 20 contiguous bases that are the same as (or the complement of) an equivalent or greater number of contiguous bases of a *Sox* gene set forth above.

- 15 The nucleic acid sequences selected as probes should be of sufficient length and sufficiently unambiguous so that false positive results are minimised. The nucleotide sequences are usually based on conserved or highly homologous nucleotide sequences or regions of SOX polypeptides. The nucleic acids used as probes may be degenerate at one or more positions. The use of degenerate oligonucleotides may be of particular
20 importance where a library is screened from a species in which preferential codon usage in that species is not known.

- Preferred regions from which to construct probes include 5' and/or 3' coding sequences, sequences predicted to encode ligand binding sites, and the like. For
25 example, either the full-length cDNA clone disclosed herein or fragments thereof can be used as probes. Preferably, nucleic acid probes of the invention are labelled with suitable label means for ready detection upon hybridisation. For example, a suitable label means is a radiolabel. The preferred method of labelling a DNA fragment is by incorporating $\alpha^{32}\text{P}$ dATP with the Klenow fragment of DNA polymerase in a random
30 priming reaction, as is well known in the art. Oligonucleotides are usually end-labelled with $\gamma^{32}\text{P}$ -labelled ATP and polynucleotide kinase. However, other methods (e.g. non-

radioactive) may also be used to label the fragment or oligonucleotide, including e.g. enzyme labelling, fluorescent labelling with suitable fluorophores and biotinylation.

After screening the library, e.g. with a portion of DNA including substantially the
5 entire Sox1-encoding sequence or a suitable oligonucleotide based on a portion of said
DNA, positive clones are identified by detecting a hybridisation signal; the identified
clones are characterised by restriction enzyme mapping and/or DNA sequence analysis,
and then examined, e.g. by comparison with the sequences set forth herein, to ascertain
10 whether they include DNA encoding a complete Sox1 (i.e., if they include translation
initiation and termination codons). If the selected clones are incomplete, they may be
used to rescreen the same or a different library to obtain overlapping clones. If the
library is genomic, then the overlapping clones may include exons and introns. If the
library is a cDNA library, then the overlapping clones will include an open reading
15 frame. In both instances, complete clones may be identified by comparison with the
DNAs and deduced amino acid sequences provided herein.

It is envisaged that SOX-encoding sequences can be readily modified by nucleotide
substitution, nucleotide deletion, nucleotide insertion or inversion of a nucleotide
stretch, and any combination thereof. Such mutants can be used e.g. to produce a
20 mutant SOX polypeptide that has an amino acid sequence differing from the sequences
of SOX polypeptides as found in nature. Mutagenesis may be predetermined (site-
specific) or random. A mutation which is not a silent mutation must not place sequences
out of reading frames and preferably will not create complementary regions that could
hybridise to produce secondary mRNA structure such as loops or hairpins.

25

Sorting of cells, based upon detection of expression of *Sox* genes, may be performed by
any technique known in the art, as exemplified above. For example, cells may be
sorted by flow cytometry or FACS. For a general reference, see Flow Cytometry and
Cell Sorting: A Laboratory Manual (1992) A. Radbruch (Ed.), Springer Laboratory,
30 New York.

Flow cytometry is a powerful method for studying and purifying cells. It has found wide application, particularly in immunology and cell biology: however, the capabilities of the FACS can be applied in many other fields of biology. The acronym F.A.C.S. stands for Fluorescence Activated Cell Sorting, and is used interchangeably with "flow cytometry". The principle of FACS is that individual cells, held in a thin stream of fluid, are passed through one or more laser beams, causing light to be scattered and fluorescent dyes to emit light at various frequencies. Photomultiplier tubes (PMT) convert light to electrical signals, which are interpreted by software to generate data about the cells. Sub-populations of cells with defined characteristics can be identified and automatically sorted from the suspension at very high purity ($\sim 100\%$).

FACS machines collect fluorescence signals in one to several channels corresponding to different laser excitation and fluorescence emission wavelengths. Fluorescent labelling allows the investigation of many aspects of cell structure and function. The most widely used application is immunofluorescence: the staining of cells with antibodies conjugated to fluorescent dyes such as fluorescein and phycoerythrin. This method is often used to label molecules on the cell surface, but antibodies can also be directed at targets within the cell. In direct immunofluorescence, an antibody to a particular molecule, the SOX polypeptide, is directly conjugated to a fluorescent dye. Cells can then be stained in one step. In indirect immunofluorescence, the primary antibody is not labelled, but a second fluorescently conjugated antibody is added which is specific for the first antibody: for example, if the anti-SOX antibody is a mouse IgG, then the second antibody could be a rat or rabbit antibody raised against mouse IgG.

FACS can be used to measure gene expression in cells transfected with recombinant DNA encoding SOX polypeptides. This can be achieved directly, by labelling of the protein product, or indirectly by using a reporter gene in the construct. Examples of reporter genes are β -galactosidase and Green Fluorescent Protein (GFP). β -galactosidase activity can be detected by FACS using fluorogenic substrates such as fluorescein digalactoside (FDG). FDG is introduced into cells by hypotonic shock, and is cleaved by the enzyme to generate a fluorescent product, which is trapped within the

cell. One enzyme can therefore generate a large amount of fluorescent product. Cells expressing GFP constructs will fluoresce without the addition of a substrate. Mutants of GFP are available which have different excitation frequencies, but which emit fluorescence in the same channel. In a two-laser FACS machine, it is possible to distinguish cells which are excited by the different lasers and therefore assay two transfections at the same time.

Alternative means of cell sorting may also be employed. For example, the invention comprises the use of nucleic acid probes complementary to Sox mRNA. Such probes can be used to identify cells expressing SOX polypeptides individually, such that they may subsequently be sorted either manually, or using FACS sorting. Nucleic acid probes complementary to Sox mRNA may be prepared according to the teaching set forth above, using the general procedures as described by Sambrook et al (1989).

In a preferred embodiment, the invention comprises the use of an antisense nucleic acid molecule, complementary to a Sox mRNA, conjugated to a fluorophore which may be used in FACS cell sorting.

Suitable imaging agents for use with FACS may be delivered to the cells by any suitable technique, including simple exposure thereto in cell culture, delivery of transiently expressing nucleic acids by viral or non-viral vector means, liposome-mediated transfer of nucleic acids or imaging agents, and the like.

The invention, in certain embodiments, includes antibodies specifically recognising and binding to SOX polypeptides. For example, such antibodies may be generated against the SOX polypeptides having the amino acid sequences set forth above. Alternatively, SOX polypeptides or fragments thereof (which may also be synthesised by *in vitro* methods) are fused (by recombinant expression or an *in vitro* peptidyl bond) to an immunogenic polypeptide and this fusion polypeptide, in turn, is used to raise antibodies against a SOX epitope.

Anti-SOX antibodies may be recovered from the serum of immunised animals. Monoclonal antibodies may be prepared from cells from immunised animals in the conventional manner.

- 5 The antibodies of the invention are useful for identifying SOX1 in neural cells expressing Sox1, in accordance with the present invention.

Antibodies according to the invention may be whole antibodies of natural classes, such as IgE and IgM antibodies, but are preferably IgG antibodies. Moreover, the invention
10 includes antibody fragments, such as Fab, F(ab')₂, Fv and ScFv. Small fragments, such as Fv and ScFv, possess advantageous properties for diagnostic and therapeutic applications on account of their small size and consequent superior tissue distribution.

The antibodies may comprise a label. Especially preferred are labels which allow the
15 imaging of the antibody in neural cells in vivo. Such labels may be radioactive labels or radioopaque labels, such as metal particles, which are readily visualisable within tissues. Moreover, they may be fluorescent labels or other labels which are visualisable in tissues and which may be used for cell sorting.

20 Recombinant DNA technology may be used to improve the antibodies of the invention. Thus, chimeric antibodies may be constructed in order to decrease the immunogenicity thereof in diagnostic or therapeutic applications. Moreover, immunogenicity may be minimised by humanising the antibodies by CDR grafting [see European Patent Application 0 239 400 (Winter)] and, optionally, framework modification.

25

Antibodies according to the invention may be obtained from animal serum, or, in the case of monoclonal antibodies or fragments thereof, produced in cell culture. Recombinant DNA technology may be used to produce the antibodies according to established procedure, in bacterial or preferably mammalian cell culture. The selected
30 cell culture system preferably secretes the antibody product.

Therefore, the present invention includes a process for the production of an antibody according to the invention comprising culturing a host, e.g. *E. coli* or a mammalian cell, which has been transformed with a hybrid vector comprising an expression cassette comprising a promoter operably linked to a first DNA sequence encoding a signal peptide linked in the proper reading frame to a second DNA sequence encoding said protein, and isolating said protein.

Multiplication of hybridoma cells or mammalian host cells in vitro is carried out in suitable culture media, which are the customary standard culture media, for example Dulbecco's Modified Eagle Medium (DMEM) or RPMI 1640 medium, optionally replenished by a mammalian serum, e.g. foetal calf serum, or trace elements and growth sustaining supplements, e.g. feeder cells such as normal mouse peritoneal exudate cells, spleen cells, bone marrow macrophages, 2-aminoethanol, insulin, transferrin, low density lipoprotein, oleic acid, or the like. Multiplication of host cells which are bacterial cells or yeast cells is likewise carried out in suitable culture media known in the art, for example for bacteria in medium LB, NZCYM, NZYM, NZM, Terrific Broth, SOB, SOC, 2 x YT, or M9 Minimal Medium, and for yeast in medium YPD, YEPD, Minimal Medium, or Complete Minimal Dropout Medium.

In vitro production provides relatively pure antibody preparations and allows scale-up to give large amounts of the desired antibodies. Techniques for bacterial cell, yeast or mammalian cell cultivation are known in the art and include homogeneous suspension culture, e.g. in an airlift reactor or in a continuous stirrer reactor, or immobilised or entrapped cell culture, e.g. in hollow fibres, microcapsules, on agarose microbeads or ceramic cartridges.

Large quantities of the desired antibodies can also be obtained by multiplying mammalian cells in vivo. For this purpose, hybridoma cells producing the desired antibodies are injected into histocompatible mammals to cause growth of antibody-producing tumours. Optionally, the animals are primed with a hydrocarbon, especially mineral oils such as pristane (tetramethyl-pentadecane), prior to the injection. After

one to three weeks, the antibodies are isolated from the body fluids of those mammals. For example, hybridoma cells obtained by fusion of suitable myeloma cells with antibody-producing spleen cells from Balb/c mice, or transfected cells derived from hybridoma cell line Sp2/0 that produce the desired antibodies are injected
5 intraperitoneally into Balb/c mice optionally pre-treated with pristane, and, after one to two weeks, ascitic fluid is taken from the animals.

The cell culture supernatants are screened for the desired antibodies, preferentially by immunofluorescent staining of cells expressing SOX polypeptides, by immunoblotting,
10 by an enzyme immunoassay, e.g. a sandwich assay or a dot-assay, or a radioimmunoassay.

For isolation of the antibodies, the immunoglobulins in the culture supernatants or in the ascitic fluid may be concentrated, e.g. by precipitation with ammonium sulphate,
15 dialysis against hygroscopic material such as polyethylene glycol, filtration through selective membranes, or the like. If necessary and/or desired, the antibodies are purified by the customary chromatography methods, for example gel filtration, ion-exchange chromatography, chromatography over DEAE-cellulose and/or (immuno-)affinity chromatography, e.g. affinity chromatography with SOX protein or with
20 Protein-A.

The invention further concerns hybridoma cells secreting the monoclonal antibodies of the invention. The preferred hybridoma cells of the invention are genetically stable, secrete monoclonal antibodies of the invention of the desired specificity and can be
25 activated from deep-frozen cultures by thawing and recloning.

The invention also concerns a process for the preparation of a hybridoma cell line secreting monoclonal antibodies directed against SOX polypeptides, characterised in that a suitable mammal, for example a Balb/c mouse, is immunised with purified SOX
30 protein, an antigenic carrier containing purified SOX polypeptide or with cells bearing SOX polypeptides, antibody-producing cells of the immunised mammal are fused with

cells of a suitable myeloma cell line, the hybrid cells obtained in the fusion are cloned, and cell clones secreting the desired antibodies are selected. For example spleen cells of Balb/c mice immunised with cells bearing SOX polypeptides are fused with cells of the myeloma cell line PAI or the myeloma cell line Sp2/0-Ag14, the obtained hybrid
5 cells are screened for secretion of the desired antibodies, and positive hybridoma cells are cloned.

Preferred is a process for the preparation of a hybridoma cell line, characterised in that Balb/c mice are immunised by injecting subcutaneously and/or intraperitoneally
10 between 10 and 10⁷ and 10⁸ cells of human tumour origin which express SOX polypeptides containing a suitable adjuvant several times, e.g. four to six times, over several months, e.g. between two and four months, and spleen cells from the immunised mice are taken two to four days after the last injection and fused with cells of the myeloma cell line PAI in the presence of a fusion promoter, preferably
15 polyethylene glycol. Preferably the myeloma cells are fused with a three- to twentyfold excess of spleen cells from the immunised mice in a solution containing about 30 % to about 50 % polyethylene glycol of a molecular weight around 4000. After the fusion the cells are expanded in suitable culture media as described hereinbefore, supplemented with a selection medium, for example HAT medium, at regular intervals
20 in order to prevent normal myeloma cells from overgrowing the desired hybridoma cells.

The invention also concerns recombinant DNAs comprising an insert coding for a heavy chain variable domain and/or for a light chain variable domain of antibodies
25 directed to the extracellular domain of SOX polypeptides as described hereinbefore. By definition such DNAs comprise coding single stranded DNAs, double stranded DNAs consisting of said coding DNAs and of complementary DNAs thereto, or these complementary (single stranded) DNAs themselves.

30 Furthermore, DNA encoding a heavy chain variable domain and/or for a light chain variable domain of antibodies directed against SOX polypeptides can be enzymatically

or chemically synthesised DNA having the authentic DNA sequence coding for a heavy chain variable domain and/or for the light chain variable domain, or a mutant thereof. A mutant of the authentic DNA is a DNA encoding a heavy chain variable domain and/or a light chain variable domain of the above-mentioned antibodies in which one or more amino acids are deleted or exchanged with one or more other amino acids. Preferably said modification(s) are outside the CDRs of the heavy chain variable domain and/or of the light chain variable domain of the antibody. Such a mutant DNA is also intended to be a silent mutant wherein one or more nucleotides are replaced by other nucleotides with the new codons coding for the same amino acid(s). Such a mutant sequence is also a degenerated sequence. Degenerated sequences are degenerated within the meaning of the genetic code in that an unlimited number of nucleotides are replaced by other nucleotides without resulting in a change of the amino acid sequence originally encoded. Such degenerated sequences may be useful due to their different restriction sites and/or frequency of particular codons which are preferred by the specific host, particularly *E. coli*, to obtain an optimal expression of the heavy chain murine variable domain and/or a light chain murine variable domain.

The term mutant is intended to include a DNA mutant obtained by in vitro mutagenesis of the authentic DNA according to methods known in the art.

20

For the assembly of complete tetrameric immunoglobulin molecules and the expression of chimeric antibodies, the recombinant DNA inserts coding for heavy and light chain variable domains are fused with the corresponding DNAs coding for heavy and light chain constant domains, then transferred into appropriate host cells, for example after incorporation into hybrid vectors.

25

The invention therefore also concerns recombinant DNAs comprising an insert coding for a heavy chain murine variable domain of an antibody directed against SOX polypeptides fused to a human constant domain γ , for example $\gamma 1$, $\gamma 2$, $\gamma 3$ or $\gamma 4$, preferably $\gamma 1$ or $\gamma 4$. Likewise the invention concerns recombinant DNAs comprising an

30

insert coding for a light chain murine variable domain of an antibody directed to SOX polypeptides fused to a human constant domain κ or λ , preferably κ .

5 In another embodiment the invention pertains to recombinant nucleic acids wherein the heavy chain variable domain and the light chain variable domain are linked by way of a DNA insert coding for a spacer group, optionally comprising a signal sequence facilitating the processing of the antibody in the host cell and/or a DNA coding for a peptide facilitating the purification of the antibody and/or a DNA coding for a cleavage site and/or a DNA coding for a peptide spacer and/or a DNA coding for an effector
10 molecule, such as a label.

According to a further aspect, and as referred to above, neuroblastic cells may be actively sorted from other cell types by detecting Sox1 expression in vivo using a reporter system. For example, such a reporter system may comprise a readily
15 identifiable marker under the control of a SOX activated expression system. Fluorescent markers, which can be detected and sorted by FACS, are preferred. Especially preferred are GFP and luciferase.

Alternatively, an *in vivo* construct expressing a reporter may be placed under the
20 control of the *Sox* control sequences themselves. These sequences are activated at the same time as *Sox* expression is activated, and therefore mark the transition into the neural pathway with the same accuracy as the *Sox* gene of interest. Advantageously, the *Sox* control sequences used are human *Sox* control sequences.

25 In general, reporter constructs useful for detecting neural cells by expression of a reporter gene may be constructed the general teaching of Sambrook et al (1989). Typically, constructs according to the invention comprise a promoter by Sox1, and a coding sequence encoding the desired reporter constructs, for example of GFP or luciferase. Vectors encoding GFP and luciferase are known in the art and available
30 commercially.

It is known that SOX proteins bind to a defined sequence motif . For example, Sox1 binds to A/T A/T CAA A/T G with high affinity. Accordingly, constructs according to the invention advantageously comprise SOX binding elements, or a functional equivalent thereof, operably linked to a gene encoding a selectable marker.

5

When transfected into cells which are potentially express SOX polypeptides, constructs according to the invention will be activated specifically by SOX polypeptide expression. Therefore, the selectable marker will be expressed once the cell enters the desired differentiation state which correlates with expression of the relevant SOX polypeptide.

10 This allows cells entering the neural differentiation pathway to be sorted by FACS.

In a still further aspect, the present invention relates to the transfection of pluripotent precursor cells, capable of differentiating into cells of a desired lineage, with a vector expressing a SOX polypeptide. By such means, pluripotent precursor cells may be
15 induced to differentiate along a desired pathway, becoming partially committed cells capable of differentiating into a variety of specialised tissues.

Herein, terms such as "transfection", "transformation" and the like are not intended to be significant, except to indicate that nucleic acid is transferred to a cell or organism in
20 functional form. Such terms include various means of transferring nucleic acids to cells, including transfection with CaPO_4 , electroporation, viral transduction, lipofection, delivery using liposomes and other delivery vehicles, biolistics and the like.

Suitable pluripotent precursor cells may be derived from a number of sources. For
25 example, ES cells, such as human ES cells and cells derived from a Germ cells (EG cells) may be derived from embryonal tissue and cultured as cell lines (Thomson *et al.*, (1998) Science 282:1145-1147). Alternatively, pluripotent cells may be prepared by a retrodifferentiation, by the administration of growth factors or otherwise, or by cloning, such as by nuclear transfer from an adult cell to a pluripotent cell such as an
30 ovum.

Human stem cells of specific lineages may be isolated from human tissues directly. Alternatively, stem cells from non-human animals, such as rodents, may be used.

Stem cells may also be propagated in vitro, for example as described in Snyder *et al.* (1996) *Clinical Neuroscience* 3: 310-316, and Martinez-Serrano *et al.*, (1996) *Clinical Neuroscience* 3:301-309. Moreover, pluripotent cell lines, such as the N-Tera II cell line which are capable of differentiating into neural cells upon stimulation with agents such as retinoic acid, are also responsive to Sox stimulation.

10 The cDNA or genomic DNA encoding native or mutant SOX polypeptides, or a label under to control of Sox sequences or a sequence transactivatable by a SOX polypeptide, can be incorporated into vectors according to techniques known in the art. As used herein, vector (or plasmid) refers to discrete elements that are used to introduce heterologous DNA into cells for expression. Selection and use of such vehicles are well
15 within the skill of the artisan. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, a transcription termination sequence and a signal sequence.

20 Most expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another class of organisms for expression. For example, a vector is cloned in *E. coli* and then the same vector is transfected into mammalian cells even though it is not capable of replicating independently of the host cell chromosome.

25

Advantageously, an expression and cloning vector may contain a selection gene, also referred to as selectable marker, other than that intended for marking Sox-expressing cells. This gene may encode a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed
30 with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and other

toxins, e.g. ampicillin, neomycin, methotrexate or tetracycline, complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.

Since the replication of vectors is conveniently done in *E. coli*, an *E. coli* genetic marker and an *E. coli* origin of replication are advantageously included. These can be obtained from *E. coli* plasmids, such as pBR322, Bluescript® vector or a pUC plasmid, e.g. pUC18 or pUC19, which contain both *E. coli* replication origin and *E. coli* genetic marker conferring resistance to antibiotics, such as ampicillin.

Expression vectors usually contain a promoter that is recognised by the host organism and is operably linked to a *Sox* gene, or label-encoding, nucleic acid. Such a promoter may be inducible by factors which induce *Sox* gene expression, or by a SOX polypeptide itself. The promoters are operably linked to DNA encoding a SOX polypeptide by removing the promoter from the source DNA and inserting the isolated promoter sequence into the vector. Both the native *Sox* promoter sequences and many heterologous promoters may be used to direct amplification and/or expression of SOX DNA. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

Control sequences, comprising a promoter and optionally enhancer(s), may be derived from the human or other *Sox* genes. Alternatively, any suitable promoter may be used, when placed under the control of a SOX-inducible element. In such a construct, the promoter selected should have a low residual level of activity, such as to minimise expression of the label in the absence of SOX polypeptide expression.

The vectors may also contain sequences necessary for the termination of transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain

nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding a SOX polypeptide or the label.

5 An expression vector includes any vector capable of expressing SOX polypeptide or label-encoding nucleic acids that are operatively linked with regulatory sequences, such as promoter regions, that are capable of expression of such DNAs. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well
10 known to those with ordinary skill in the art and include those that are replicable in eukaryotic and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome. For example, DNAs encoding SOX1 may be inserted into a vector suitable for expression of cDNAs in mammalian cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, et al., (1989) NAR 17, 6418).

15 Particularly useful for practising the present invention are expression vectors that provide for the transient expression of DNA encoding a SOX polypeptide or a label in mammalian cells. Transient expression usually involves the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates
20 many copies of the expression vector, and, in turn, synthesises high levels of the SOX polypeptide or a label. For the purposes of the present invention, transient expression systems are useful e.g. for identifying SOX expressing cells or for inducing a pluripotent cell to differentiate.

25 Construction of vectors according to the invention employs conventional techniques, for example as described in Sambrook *et al.*, 1989. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors,
30 preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing gene expression and function are known to those skilled in the art. Gene

presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridisation, using an appropriately labelled probe which may be based on a sequence provided
 5 herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

The invention is described, for the purpose of illustration only, in the following examples.

10

MATERIAL AND METHODS

Manufacture of SOX1 polyclonal antibodies: A 622bp HincII fragment encoding sequences C-terminal of the HMG box of SOX1 (207 a.a.) is fused in frame to the bacterial GST gene in the construct pGEX3X. Fusion protein is induced and purified
 15 as described by Smith and Johnson (1988). rabbits are treated with a course of injections as recommended by Smith and Johnson (1988): each injection contains 250µ of fusion protein. Two final bleeds, FB43 and FB44 , are obtained from the rabbits prior to the preparation of polyclonal sera.

20 **Immunocytochemistry:** Embryos, P19 cells and neural plate explants are examined using standard techniques (Placzek *et al.*, 1993). Antibodies are used at the following dilutions: anti-SOX1 PAb (1:500); K2 anti-HNF3β MAb (1:40); 6G3 anti-FP3 MAb (1:10); anti-3A10 MAb (1:10); anti-2H3(Neurofilament-160) MAb (1:10); 4D5 anti-Islet-1 MAb (1:1000); anti-SSEA1 MAb (1:80) (Hybridoma Bank); anti-NESTINE
 25 MAb (1:10) (Hybridoma Bank); anti-BrDU MAb (1:500) (Sigma); Appropriate secondary antibodies (TAGO and Sigma) are conjugated to fluorescein isothiocyanate (FITC), Cy2 or Cy3.

BrDU analysis: Pregnant mice are injected intraperitoneally with 50µg/g of body weight of 5-bromo-2deoxyuridine (BrDU) (Sigma) in 0.9% NaCl and sacrificed two
 30 hours after injection. Embryos are fixed and sectioned as described above. The slides

are washed twice in PBS, and incubated in 0.2% HCl at 37°C for 30 minutes, then rinsed thoroughly with PBS, followed by three rinses with PBS/0.1% Triton/1% heat inactivated goat serum (P-T-G). Monoclonal anti-BrDU (1:500 dilution in P-T-G) is applied to the sections and incubated at 4°C overnight. Sequential sections are
5 incubated in SOX1 antibody (1:500 dilution in P-T-G) at 4°C overnight. The slides are washed twice in P-T-G, then incubated in the appropriate secondary antibody for 30 minutes at room temperature, washed with P-T-G and mounted.

P19 cell cultured and retinoic acid treatment: P19 cells are cultured as previously
10 described (Rudnichy and McBurney, 1987). To induce differentiation, cells are allowed to aggregate in bacterial grade petri dishes alone, in the presence of 1µM retinoic acid or in the presence of 1µM retinoic acid or in the presence of 5mM IPTG. After 4 days of aggregation in the presence of inducing agents, cells are plated on tissue culture chamber slides. The cells are allowed to adhere and grow for 4-5 days, with
15 media changes every 24 hours. For immunofluorescence, cells are grown on tissue culture chamber slides coated with 0.1% gelatin, washed once with PBS, fixed at room temperature in 1x MEMFA for 1 hour, washed in P-T-G twice; then stained with appropriate antibody.

Cell counting analysis: For cell counting experiments P19 transfectant cell lines are
20 induced to differentiate, plated on gelatine coated slides, fixed at room temperature in 1xMEMFA for one hour at day 6-8 for neurons. Cells are stained with Neurofilament (2H3) antibody and photographed using an Olympus fluorescence microscope. Cell counts are expressed as percentages of total cells in a field. Eight fields from two
25 different experiments are counted for each P19 clone.

Plasmids and transfection: To construct the SOX1 expression vector, pRSVopSox1, the POP113CAT operator vector (Stratagene) is digested with Not1, end-filled Kpn/Stu (position 431-1694) fragment of the *Sox1* cDNA. The P3'SS, eukaryotic Lac repressor
30 expressing vector (obtained from Stratagene) is transfected into P19 cells by lipofection. Stable transformants are selected in 250 µg/ml of hygromycin. Expanded clones (250)

are isolated and examined for expression of the Lac repressor by indirect immunofluorescence with anti-lac PAb (Stratagene). Four cell lines are isolated (P3'SS-10, 13, 22 and 47) which show ubiquitous and constitutive expression of the Lac repressor. P3'SS-10 is chosen for the subsequent experiments. P3'SS-10 is then
 5 transfected with pRSVopSox1 by lipofection. Stable clones are selected using 500µg/ml G481. 250 clones are expanded and analysed for inducible *Sox1* expression by RNase protection and immunocytochemistry with SOX1 antibody.

RNase protection assays: Total RNA is prepared from P19 cells and RNase protection
 10 assays are carried out using 5µg of P19 cell RAN as described by Capel *et al.*, (1993). Anti-sense labelled probes are derived from the 396 bp SmaI-BspH1 fragment (position 1467-1863) of the *Sox1* cDNA, a 215bp BsaI exon 4 specific fragment of *Wnt1* cDNA, a PvuII digest of the *Mash1* cDNA (Johnson *et al.*, 1992) and a NotI digest of SAP D cDNA is used a loading control (Dresser *et al.*, 1995).

15

RT-PCR: Total RNA is prepared from P19 cells as described by Capel *et al.*, (1993). Reverse transcription, PCR reaction, and primers is performed as described by Okabe *et al.*, (1996)

Rat lateral neural plate explants: Lateral neural plates (LNP) are isolated from days
 20 8.5-9.0 rat embryos from prospective hindbrain and spinal cord regions as previously described (Placzek *et al.*, 1993). Notochord explants are dissected from HH stage 608 chick embryos as previously described (Placzek *et al.*, 1993). Explants are embedded in collagen and cultured (Placzek *et al.*, 1993) for 24, 48 and 96 hours. Purified rat
 25 SHH-N (Ericson) *et al.*, 1996) is added to cultures at concentrations within the effective ranges used in other assays (Ericson *et al.*, 1996)

EXAMPLE 1

SOX1 IS EXPRESSED DURING EARLY NEURAL DEVELOPMENT

30

SOX1 expression during mouse and rat neurulation is analysed using a rabbit polyclonal antibody against the SOX1 C-terminal region. In the mouse, expression of SOX1 is first detected at 7.5 days post coitum (*dpc*) in the anterior half of the late-streak egg cylinder. Cross-sections through the embryo at this stage reveal expression in columnar
 5 ectodermal cells, which appear to define the neural plate, while cells located more laterally are negative. Thus, SOX1 expression at this stage is specific to the neural plate. SOX1 is maintained in all neuroepithelial cells along the entire anteroposterior axis as the neural plate bends (8.0-8.5 *dpc*, as shown in cross-sections of a 2 somite mouse embryos where Sox1 expression is limited to neural folds) and fuses to form the
 10 neural tube (9.0-9.5 *dpc*, where Sox1 labelling is seen to be restricted to the neural tube in cross-sections of 10-12 somite mouse embryos). The pattern of expression of SOX1 in the rat is similar to that in the mouse. The expression of SOX1 throughout the neural plate and early neural tube implies a similarity amongst these cells.

15 After neural tube closure, neuroepithelial cells begin to differentiate into defined classes of neurons at specific dorsoventral (D/V) positions within the spinal cord (Altman and Bayer 1984, Tanabe and Jessell, 1996). As development proceeds, Sox1 is downregulated in a stereotyped manner in cells along D/V axis of the neural tube. In the spinal cord, expressions first downregulated in cells that occupy the ventral midline
 20 (cross-sections of the thoracic region of 20 somite mouse embryos reveal a lack of SOX1 staining in this area), then the ventral motor horns (corresponding lack of staining being visible in cross section of 30-35 somite embryos) and subsequently the dorsal regions. These regions appear to correlate with floor plate, motor neurons and sensory relay interneurons, respectively.

25

To ascertain this a series of antibody double-labelling experiments are performed in rat embryos. The SOX1 antibody is used in combination with a panel of antigenic markers which identify cells of the floor plate and mature neurons (Neurofilament (NF-1): labelled with contrasting colour markers and visualised in an E11 rat embryo).
 30 Expression of SOX1 and expression of these markers is almost entirely mutually exclusive. In the ventral spinal cord or the 10.0-12.0 *dpc* mouse embryo, SOX1

expression is maintained only in 'region X' (Yamada *et al.*, 1991), as revealed by immunolabelling of two streams of cells located between the differentiated floor plate and ventral motor horns in 30-35 somite embryos. Eventually, by 13.5 *dpc*, SOX1 expression is restricted to a thin ventricular zone in the CNS. SOX1 expression is detected in the peripheral nervous system (PNS). These expression profiles suggest that SOX1 is expressed by early neural cells in the CNS and is downregulated in the developing neural tube coincident with neural differentiation.

EXAMPLE 2

10 SOX1 MARKS PROLIFERATION CELLS WITHIN THE EMBRYONIC NEURAL TUBE

The uniform expression of SOX1 in the neural plate and early neural tube followed by its down regulation along the D/V axis and restriction to the ventricular zone is reminiscent of the pattern of cell proliferation in the developing central nervous system (Sauer, 1935; Fujita, 1963; Altman and Bayer, 1984). In the neural plate and early neural tube, proliferating progenitor cells are organised in a pseudostratified epithelium in which the processes of these cells extend from the inner luminal to the outer mantle surface. At later stages the neural tube becomes progressively thicker and can be divided into different zones. The proliferating CNS progenitors are largely restricted to the inner ventricular zone (VZ) around the lumen. They begin to migrate away from the lumen while in S-phase, and after completing their final mitosis, migrate to the outer layer, the marginal zone (MZ). In the 10.5 *dpc* mouse embryo, SOX1 expression is detected, using an anti-SOX1 antibody, throughout the pseudostratified epithelium of the posterior neural tube and is restricted to the ventricular zone in more mature anterior region of the neural tube. In order to evaluate the relationship between SOX1 expression and proliferating CNS cells are directly assayed proliferation by monitoring the incorporation of bromodeoxyuridine (BrDU) with an anti-BrDU antibody. Pregnant mouse females at 10.5 *dpc* are injected with BrDU two hours prior to dissection to detect proliferating cells. Embryos are then fixed, sectioned and double-labelled for BrDU incorporation and SOX1 expression. Similar to SOX1 expressing cells, those

that incorporate BrDU are found throughout the posterior neural tube in 10.5 *dpc* mouse embryos and lie in the ventricular zone of the anterior neural tube. All cells that incorporate BrDU also express SOX1. SOX1-positive cells that do not incorporate BrDU are restricted to the luminal surface of the ventricular zone. In contrast, no
 5 SOX1 nor BrDU-positive cells are detected in the outer marginal zone. These results show that SOX1 is expressed in dividing neuroepithelial cells within the embryonic CNS.

EXAMPLE 3

10 SOX1 IS DOWNREGULATED IN COMMITTED CELLS

The mutual exclusion of SOX1 and markers of committed differentiated cells such as Islet1 (Pfaff *et al.*, 1996) raises the possibility that the downregulation of SOX1 may be a pre-requisite step for the differentiation in neural plate explants *in vitro*. Isolated
 15 neural plates explants are cultured with known inducers of ventral neural cells, namely the notochord and purified Sonic Hedgehog protein. The expression of SOX1 and incorporation of BrDU is then compared to the expression of three markers of ventral cells, Islet1, FP3 and HNF3 β . Consistent with our observations *in vivo* both the expression of SOX1 and Islet1 as well as SOX1 and FP3 is mutually exclusive in neural
 20 plate explants cultured adjacent to notochord (n=8) or in the presence of purified Sonic Hedgehog protein as seen in E9 rat neural plate tissue cultured with Sonic Hedgehog protein for 48 hours and stained with anti-SOX1 and anti-Islet1 antibodies. Similarly, the incorporation of both BrDU and Islet1 as well as BrDU and FP3 (detected using an anti-FP3 antibody) is mutually exclusive. In contrast, the domain of expression of
 25 HNF3 β is found to extend beyond that of FP3 and into the region of BrDU positive cells.

To determine whether a similar population of cells could be detected *in vivo* embryos are analysed, and for co-expression of FP3 and HNF3 β and for co-expression of BrDU
 30 and HNF3 β . We find that medial floor plate cells co-express HNF3 β and FP3 but do not incorporate BrDU, whereas lateral floor plate cells express only HNF3 β and

incorporate BrDU. HNF3 β thus provides a marker for cells that are mitotically active but have begun to differentiate.

These cells, occupying the medial regions of the floor plate, express HNF3 β but not SOX1. In contrast cells occupying lateral regions of the floor plate co-express HNF3 β and SOX1. These observations, together with the mutually exclusive expression of SOX1 with Islet 1 and FP3 in ventral neural cells provide evidence that SOX1 is downregulated as cells exit mitosis and not at the onset of cell differentiation.

10 **EXAMPLE 4**

SOX1 EXPRESSION IS ASSOCIATED WITH NEURAL DIFFERENTIATION

Neural induction is accompanied by the onset of new gene expression which in turn enables the formation of neural rather than epidermal tissue. The early and apparently uniform expression of SOX1 in neural cells, together with observations that Sox genes may affect cell lineage decisions (see Introduction), raises the possibility that SOX1 expression is an early response to neural inducing signals and that its expression may be involved in directing cells towards a neural fate. To address whether SOX1 plays a role in establishing neural fate in response to A P19 cell culture system is used as an *in vivo* model system in which to analyse SOX1 expression and the effects of its misexpression.

P19 cells are an embryonal carcinoma cell line with the ability to differentiate into all three germ layers (McBurney, 1993). In the undifferentiated state P19 cells morphologically resemble an uncommitted primitive ectodermal cell and express the cell surface antigen SSEA-1. These cells have a very low rate of spontaneous differentiation when grown in a monolayer in the absence of chemical inducers. P19 cells grown as aggregates, however, differentiate partially into endodermal cells. Furthermore, with the addition of retinoic acid, aggregated P19 cells differentiate into neuroepithelial-like cells (Jone-Villeneuve *et al.*, 1982). These express neuroepithelial markers such as NCAM, intermediate filament NESRIN, MASHI (Johnson *et al.*,

1992) and WNT1 (St. Arnaud *et al.*, 1989). When plated onto a substrate, about 15% of these cells differentiate into mature neurons expressing Neurofilament. Thus, in this *in vitro* model system retinoic acid acts as a “neural inducer”.

- 5 Initially, the expression of *Sox1* in P19 cells is examined by both RNase protection and immunocytochemistry. The features of *Sox1* expression in P19 cells are similar to those observed in prospective neural tissue *in vivo*. *Sox1* mRNA and protein can not be detected in undifferentiated P19 cells which express the cell-surface antigen SSEA1 when analysed using anti-SOX1 and anti-SSEA antibodies, and by RNase protection.
- 10 Similarly, when P19 cells are differentiated as aggregates without the addition of chemical inducers, SOX1 is not expressed as determined by RNase protection. In contrast, SOX1 is rapidly induced during neural differentiation when aggregated P19 cells are differentiated in the presence of retinoic acid. *Sox1* thus behaves similarly to other neuroepithelial markers such as *Mash 1* and *Wnt 1*, the transcripts of which are
- 15 detected in retinoic acid-treated P19 cells by RNase protection.

When retinoic acid-treated P19 cell aggregates are plated onto tissue culture substrate, about 15% of the cells differentiate into mature process-bearing, Neurofilament-expressing neurons. Double-label immunofluorescence is used to simultaneously detect

20 SOX1 and Neurofilament, to examine the expression of SOX1 in P19 cells displaying a fully differentiated neuronal morphology. SOX1 immunoreactivity is not detected in process-bearing Neurofilament-positive neurons. Thus, as *in vivo*, SOX1 is expressed by P19 cells when they first assume a neural fate but it is then downregulated with their differentiation.

25

EXAMPLE 5

USE OF SOX1 TO DIRECT CELLS TO A NEURAL FATE

The previous data suggest that in P19 cells, as *in vivo*, SOX1 expression is induced at a

30 time when neuroepithelial cells begin to differentiate. If SOX1 plays a role in directing cells towards the neural fate, expression of SOX1 in P19 cells may be able to substitute

for retinoic acid to initiate neural differentiation. Endogenous SOX1 is accordingly activated in P19 cells using an inducible eukaryotic lac repressor-operator expression system. To establish this system a clonal line of P19 cells is generated which constitutively and ubiquitously expresses the lac repressor. This parent line (P3'SS-10) is transfected with pRSVopSox1, a vector containing the *Sox1* cDNA under the regulation of an inducible RSV promoter and stable lines are established. In the uninduced state, without the addition of isopropyl- β -d-thiogalactase (IPTG) these lines express high levels of the lac repressor that binds to operon sites upstream of the RSV promoter and thus blocks transcription of *Sox1*. Upon addition of IPTG a conformational change occurs, decreasing the affinity of the repressor and resulting in the activation of pRSVopSox1. Approximately 250 clones of transfectants are isolated in the repressed state. Using RNase protection and immunocytochemistry assays three clones are selected (708-13, 708-16 and 708-21) that express high levels of RSVopSox1 in response to IPTG.

15

The pluripotentiality of these clones is not compromised by the transfection and selection. All three lines express SSEA1 in the uninduced state. Furthermore, when aggregated in retinoic acid the uninduced clones initiate expression of endogenous *Sox1* and differentiate into mature Neurofilament-expressing neurons after plating, in a manner similar to wild-type P19 untransfected cells.

20

In order to address whether expression of SOX1 can initiate neural differentiation and thereby substitute for the requirement of retinoic acid, it is determined whether the transient exposure of P19 aggregates to retinoic acid can be replaced by a transient induction of RSVopSox1, through addition of IPTG. Wildtype P19 cells and transfected P19 clones (708-13, 708-16 and 708-21) are cultured as aggregates for 96 hours with or without the addition of IPTG. After 96 hours RNA is isolated from half of the aggregates for RNase protection and/or RT-PCR assays. The remaining aggregates are plated onto tissue culture substrate, allowed to differentiate for three days without further addition of IPTG and then scored for the expression of a panel of neuroepithelial and neuronal markers by immunocytochemistry. These conditions are

30

the same as those used for retinoic acid-induced differentiation of wildtype P19 cells. After 96 hours the clones induced to express RSVopSox1 with IPTG express endogenous *Sox1* and *Mash1*. The expression of these two neuroepithelial markers is similar to that seen in wildtype cells induced with retinoic acid. In addition the IPTG induced clones expressed NESTIN and *Hoxa7* (Mahn et al., 1988). Further differentiation of the transiently-induced clones on substrate showed the presence of mature neurons as demonstrated by Neurofilament-positive, 3A10-positive and Islet1-positive cells. All three clones 708-13, 708-16 and 708-21 differentiate in this matter although the number of mature neurons produced is variable. The number of differentiated neurons formed in the IPTG induced clones is estimated by determining the number of Neurofilament-positive cells in a given field of cells. The number of neurons ranges from 6-8% for clone 708-13, 15-20% for clone 708-16 and 20-25% for clone 708-21. The latter two clones show uniform and ubiquitous induction of SOX1 expression whereas expression in clone 708-13 is not in all cells. In addition, the transiently induced clones generate GFAP-positive cells indicating glial cell differentiation. None of these markers is detected in wildtype P19 cells cultured in the presence of IPTG or in clones 708-13, 708-16, and 708-21 cultured in the absence of IPTG. The expression of SOX1, both in vivo and in vitro, is mutually exclusive with mature neuronal markers such as Neurofilament and Islet1. To examine SOX1 expression in the mature neurons generated in the transiently-induced clones, double-label immunofluorescence is used to simultaneously detect SOX1 and Neurofilament. No SOX1 expression could be detected in cells positive for Neurofilament in these cultures.

25 **EXAMPLE 6**

USE OF SOX2 TO ISOLATE NEURAL PRECURSORS

Sox1 is confined to the neuroepithelium of the neural plate and dividing neural progenitors in the early mouse embryo, whereas Sox2 is found in an overlapping pattern that also encompasses floor plate and early neural crest cells. Thus, although Sox1 is clearly indicated for neural cell selection, the role of Sox2 is less certain.

For induction of neural differentiation, ES cells are aggregated in suspension to form embryoid bodies, exposed to retinoic acid, and then allowed to reattach to a substratum. Neuronal-like cells can be detected in the out-growths, accompanied by a variety of other cell types. Two variations are introduced to the protocol that enhances the final representation of neuronal cells. First, the embryoid bodies are dissociated before plating. This results in a homogeneous dispersion and terminates inductive and selective effects within the embryoid bodies. Second, cells are plated in a defined culture medium - DMEM/F12 plus N2 supplement - on substrata coated with poly-D-lysine and laminin, which support attachment and outgrowth of neuronal cells.

These procedures have an additive effect on the proportion of neural cells in the cultures. When combined, up to 50% of viable cells extended neuritic processes and become immunoreactive for the neuronal markers neurofilament light and heavy chains, microtubule-associated proteins, MAP2 and tau, or β -tubulin III.

Immunostaining of freshly plated cells with antibodies against Sox1 and Sox2 reveals that 40–50% of the cells are positive for each marker. This approximates to the final proportion of differentiated neural cells, consistent with the notion that cells expressing Sox1 and Sox2 correspond to neural-restricted progenitors.

To attempt to isolate the neural progenitor pool, ES cells are used in which the bifunctional selection marker/reporter gene β geo has been integrated into the *Sox2* gene by homologous recombination. When induced to differentiate as described above, approximately 50% of these cells stain for β -galactosidase activity, consistent with the proportion of cells that express Sox2 protein. Therefore, application of G418 to the differentiating cultures should eliminate Sox2-negative non-neural cells. G418 (200 g/ml) is added after retinoic-acid induction, either during embryoid body culture or upon plating. In both conditions appreciable cell killing is evident. Crucially, however, large numbers of cells survive that exhibit the small, ovoid morphology typical of neuroepithelial cells. Over 90% of these cells show prominent β -galactosidase staining.

Expression of Sox1 and Sox2 proteins is confirmed by immunostaining. Consistent with a neuroepithelial identity, the cells also express nestin.

Accordingly, neural cell types may be isolated by expression of a marker associated with Sox2, starting with a population of totipotent cells which has been induced to differentiate *inter alia* into a neural pathway.

In order to determine whether the Sox2-selected population have proliferative capacity, β FGF is added to plated cultures. This results in a major stimulation of cell division.

10 The expanded cells predominantly retain undifferentiated morphology and show strong X-gal staining indicative of Sox2 expression. Such cultures can be amplified and serially passaged for at least three weeks, which is significantly longer than the proliferative phase of neurogenesis in the mouse embryo.

15 In the absence of mitogen, Sox2-selected precursor cells begin to extend neuritic processes within 48 hours and by 96 hours form a network of neuron-like cells. The pan-neuronal markers neurofilament light chain, microtubule-associated proteins, MAP2 and tau, and β -tubulin III are detectable from 48 hours onwards, coincident with down-regulation of Sox2 expression. By 96 hours, over 90% of cells express neuronal

20 markers, including neurofilament heavy chain and synapsin I. Cells of non-neuronal morphology are rarely apparent, with the exception of the occasional GFAP-positive astrocyte. Astrocyte numbers increase if serum or FGF is added to the cultures. Maturation of the neuronal cells, evidenced by production of gamma-aminobutyric acid (GABA) and glutamate neurotransmitters, and further elongation of neurites with

25 dendritic sprouting is achieved on transfer to Neurobasal medium supplemented with B27 and horse serum.

This ability to generate pure populations of neurons, combined with the relative ease of genetic modification of ES cells, offers a new route for manipulation and

30 characterisation of neuronal development and cell biology. The finding that major cellular components of embryoid bodies can be ablated without apparently perturbing

development of the surviving cells also indicates that this strategy can be adapted to isolate stem or precursor cells for other lineages. An important attribute is that unlike immunopurification techniques this approach is not limited to cell-surface antigens but can be applied to any *Sox* gene. Selected populations can readily be refined by
5 introducing independent markers into more than one gene.

The advantage of targeting progenitors as opposed to differentiated cells is the potential for subsequent amplification and directed differentiation both *in vitro* and *in vivo*. ES cell derivatives can colonise host tissue and differentiate after transplantation into adult
10 recipients. Grafts of whole embryoid body cultures, however, also give rise to teratomas and other benign or malignant growths. Furthermore, heterologous cells may interfere with trophic signals and guidance cues from host tissue to transplanted cells. Prior lineage purification should eliminate these problems and enable the multipotentiality of ES cells to be harnessed effectively for application in cellular
15 transplantation.

Claims

1. A method for isolating a pluripotent cell which is at least partially committed to a given developmental pathway, comprising the steps of :
 - 5 a) selecting a population of pluripotent cells;
 - b) sorting the cells according to *Sox* gene expression; and
 - c) isolating those cells which express a given *Sox* gene.
- 10 2. A method according to claim 1, wherein the population of cells for is derived from CNS tissue.
3. A method according to claim 1, wherein the population of cells is derived from a cell culture.
- 15 4. A method according to any preceding claim, wherein the expression of the *Sox* gene is detected by nucleic acid hybridisation.
5. A method according to any one of claims 1 up to 3, wherein the expression of the *Sox* gene is detected by a binding of a SOX polypeptide to a detectable ligand.
- 20 6. A method according to claim 5, wherein the detectable ligand is a labelled immunoglobulin.
- 25 7. A method according to claim 5, wherein the detectable ligand is a labelled oligonucleotide complementary to *Sox* mRNA.
8. A method according to any preceding claim, wherein the expression of the *Sox* gene is detected by FACS analysis.

9. A method for isolating a desired cell type from a population of cells, comprising the steps of:
- (a) transfecting the population of cells with a genetic construct comprising a coding sequence encoding a detectable marker operatively linked to control regions sensitive to modulation by a SOX polypeptide;
 - (b) detecting the cells which express the selectable marker; and
 - (c) sorting the cells which express the selectable marker from the population of cells.
10. A method for isolating a neuroblastic cell from a population of cells, comprising the steps of:
- (a) transfecting the population of cells with a genetic construct comprising a coding sequence encoding a detectable marker operatively linked to a control sequence which is transactivatable by a SOX polypeptide;
 - (b) detecting the cells which express the selectable marker; and
 - (c) sorting the cells which express the selectable marker from the population of cells.
11. A method according to claim 9 or claim 10, wherein the selectable marker is a fluorescent or luminescent polypeptide.
12. A method according to claim 9 or claim 10, wherein the selectable marker is a polypeptide detectable at the surface of the cell.
13. A method for producing a cell committed to a specified lineage, comprising the steps of:
- (a) transfecting a pluripotent stem cell with a genetic construct comprising a coding sequence expressing a SOX polypeptide;
 - (b) culturing the stem cells in order to differentiate them into neural cells; and
 - (c) isolating the neural cells thereby produced.

14. A method according to claim 15, wherein the *Sox* sequence is operatively linked to an inducible promoter.

15. A method according to claim 13 or claim 14, wherein the cell is further
5 transfected with a vector comprising a sequence encoding a regulator which modulates the expression of the *Sox* sequence.

16. A method according to any preceding claim, wherein the *Sox* gene is a member of Sox Group A.

10

17. A method according to claim 16, wherein the *Sox* gene is *Sox1* or *Sox2*.

ABSTRACT

Sox gene expression correlates in general with specific stages during embryogenesis. It has been determined that the expression of *Sox* genes may be used, as directed herein, to induce or select pluripotent cells which are at least partially committed to a given developmental pathway. There is provided a method for isolating a pluripotent cell which is at least partially committed to a given developmental pathway, comprising the steps of : a) selecting a population of pluripotent cells; b) sorting the cells according to *Sox* gene expression; and c) isolating those cells which express a given *Sox* gene.

